# APPENDIX C-5
# Scripted Online User Assessment Methodology

# APPENDIX C-5
# Scripted Online User Assessment Methodology

## Table of Contents

# APPENDIX C-5
# SCRIPTED ONLINE USER ASSESSMENT METHODOLOGY

## 1.0. INTRODUCTION

The methods used in this research activity served four purposes. First, the online sessions permitted in-process, "front-line," collection of data concerning user assessments of GILS—as opposed to "*re*collection" of assessments after GILS use. Second, the method was geared toward eliciting highly qualitative responses to a *concept* (i.e., rather than the more traditional aims of user assessments such as quantification of relevant "hits" or usage patterns). Third, the assessment questions attempted to gain insight into cognitive processes of users in the online, networked environment. Last, investigators documented this new exploratory technique and instrument with an expectation that future research will build upon and improve them. Investigators consider these objectives paramount in understanding user perceptions, expectations, and behavior during networked information discovery and retrieval (NIDR), and in advancing the quality of GILS accordingly.

## 2.0. METHOD OVERVIEW

Graduate and undergraduate student "users" unfamiliar with GILS were oriented to the nature and purpose of their participation by means of a 5-minute verbal introduction by the investigators (see Attachment 1). They were subsequently asked to record answers to more than 50 multiple-choice, free-form expression, and true/false questions as they navigated "real life," "real time" Government Printing Office (GPO) GPO Access and Environmental Protection Agency (EPA) GILS systems according to a scripted set of encounters. The script was based on results of the record content analysis and investigators' ongoing search/retrieval experience with various GILS. The questions were designed to elicit user feedback concerning GILS content and service expectations, record design, orientation in information space, adaptation to the metadata construct (e.g., searching reflexes), and, perhaps most importantly, users' assumptions about GILS—all on the basis of this 1-hour first-exposure to scripted transactions. In addition, investigators conducted debriefing sessions where users were informed generally of GILS scope and purpose and asked to elaborate on intellectual and emotional impressions created by the scripted. The qualitative data from the sessions were entered into a database to facilitate disclosure of patterns related to users' reactions to GILS as a service concept and to GILS product. As with the record content analysis (see Appendix C-5 Record Content Analysis Methodology), investigators recorded suggested improvements to the development and execution techniques for scripted online-user assessment in order to optimize recommendations to agencies interested in adopting the techniques.

The following paragraphs present information concerning the scripted online user assessment research objectives, its context within the overall evaluation framework, data collection/analysis, data collection instrument (script), participants, session delivery, method limitations, lessons learned, and recommendations.

## 3.0. OBJECTIVES

The scripted online user assessments were conducted to capture users':

- Subjective appraisal of GILS efficacy as a tool for NIDR, including
  - Appreciation of/adaptation to the construct of metadata in the virtual environment
  - Record display/presentation
- Emotional and intellectual reactions to GILS products and services, including
  - Confidence in using GILS
  - Expectations for content and service quality
  - Perceptions of GILS as an information space
- Assumptions about GILS based on a limited first-exposure, including responses to

- Objectives
- Architecture
- Coverage
- Potential.

These objectives were realized in the data collection and analysis activities outlined below.

## 4.0. CONTEXT WITHIN EVALUATION FRAMEWORK

Both the development of the user assessment script (see Appendix D-5 Scripted Online User Assessment Instrument) and lines of inquiry during the debriefing session were informed by results of previous data collection activities. The content analysis of GILS records (see Appendix E-2 Record Content Analysis Findings, Discussion, and Recommendations) proved particularly useful in identifying the type of GILS encounters most likely to provoke user response. In addition, the content analysis provided a basis for representing the content and quality variations in GILS at macro- and micro levels (i.e., as a government-wide collection and as individual records).

The session provided a degree of triangulation against data collected by way of focus groups and the conference survey. "Users" in the focus groups fell into broad stakeholder categories of agency staff (e.g., records managers and public information officers), public interest groups (perhaps focused on access policy and scope of coverage as much if not more that actual service satisfaction), and professional intermediaries (e.g., librarians). The "users" in the scripted online assessment, however, could be considered "closer to the street" or more representative of "public access"/average citizen end-use requirements.

In addition, as discussed in Section 10 Lessons Learned and Recommendations, the methodology is presented as one in a complementary suite of activities offering benefit in the areas of system requirements definition and quality assurance.

## 5.0. DATA COLLECTION AND ANALYSIS

The objectives described in Section 3 translated into the following categories of data collection and analysis:

- Content expectations
  - Full-text of documents vs. metadata
  - Subject matter and resource types (objects represented)
  - Information quality (e.g., currency and completeness/level of detail)
  - Scope and extent of collection
  - Record and resource aggregation ("distance" from satisfaction of an information need)
- Service expectations
  - Predictability of results
  - Effectiveness of fielded searching
  - Cause of search logic and "service" errors
  - System/service response time (transaction and record download time)
  - "Comfortability" and satisfaction with GILS purpose and its translation into service
  - Quantity and nature of hyperlinks
  - Implementation policies to yield consistency across GILS systems and other objectives
- Satisfaction with record characteristics
  - Cosmetic appearance
  - Ergonomics (e.g., element display order)
  - Length
  - File formats (ASCII vs. HTML)
- Perceptions of information space and ownership

  - Navigation
  - Centralization (loci of services and products)
  - Implementation architecture (network distribution)
  - Availability of resources
  - Authority (hierarchy of service and product responsibility for quality)
- GILS nomenclature (use of bibliographical and NIDR terminology in element names and definitions)
- Searching reflexes and relevance judgments
  - Preferences for full-record vs. fielded searching
  - Relevance improvement tactics
  - Requirements for user "sophistication" *vis-a-vis* education and training

These categories, while not of equal weight or under equal levels of control by GILS implementors, compose a backdrop useful for evaluating GILS from the perspective of a capable but "GILS-unaware" online searcher or browser.

## 6.0. DESCRIPTION OF DATA COLLECTION INSTRUMENT (SCRIPT)

Appendix D-5 Scripted Online User Assessment Instrument presents the instrument used during this research activity. The first 10 items plus 3 on the last page captured demographic and other information about the participants—such as status (e.g., student, private citizen, etc.) and government-information search frequency, methods, resource types, and knowledge/attitudes about government information--provided a context for evaluating expectations and responses.

The content and organization of the script were planned to provide these first-time users with sufficient grounding to execute searching and the widest possible range of GILS experience within a prescribed 1-hour timeframe—a very difficult undertaking—especially given our methodological objective of minimizing effects of user interface and search engine functionality. Users were not permitted to hyperlink to nonscripted help information of any kind nor allowed to repeat searches in order to understand or improve results. (The reasonableness of the former constraint was confirmed during pretest of the user profile sheet below, which indicated that student users are unlikely to consult online help but rather rely on trial-and-error in searching.)

The assessment script directed users to encounter a minimum of two GILS implementations—GPO Access and EPA—and up to ten GILS sites. (Users were provided approximately 10 minutes of nondirected exploration of FedWorld, NTSB, USDA, DOL, GSA, and SBA sites.) Within the two mandatory sites, the script called for six searches (executions of both self-directed and designed query) and subsequent examination of results pages, and examination of four individual records. GPO Access and EPA were selected from among the field of hosts because they afforded an "integrated" vs. "standalone" perspectives (i.e., cross-agency vs. agency-centric models), they are relatively mature implementations, and because users were likely to be familiar with these agencies' missions. The other "optional" sites listed above were selected on the basis of their wide variety of mission, implementation approach (i.e., WAIS vs. HTTP), and types of resources described (e.g., technical reports, databases, consumer information, regulation digests, etc.).

The script attempted to provide users a "fair" or reasonable simulation of personal nondirected use while necessarily constraining movement both to control outcomes, increase cross-user data validity, and save time. With this in mind, the following text broadly outlines the script activities and encounter types in script order:

GPO Access

- Placement in GPO Access GILS space
- Exposure to GILS nomenclature
- Selection of metadata elements (search fields) and formulation of query for a scripted and broad information need
- Assessment of and reaction to search result
- Scripted retrieval of a particular record for comparison with another describing a like resource

- Scripted retrieval of two records having the same title
- Selection of metadata element (search fields), selection of GILS database, and formulation of query for a scripted and very narrow/specific information need (known-item search)
- Assessment of and reaction to search result
- Scripted retrieval of two records having the same control identifier and titles in different languages

EPA

- Placement in EPA GILS space and perceptions relative to GPO Access GILS space/function
- Enumeration of expected EPA GILS content
- Exposure to GILS nomenclature
- Scripted retrieval and subsequent assessment of a particular record's descriptiveness
- Enumeration of assumptions related to authority, extent of coverage, and availability

In addition to these active encounters, the script included a question about users' assumption of the GILS universe extent and 14 post-searching Likert scale (5 points: strongly agree↔strongly disagree, no opinion) questions about the desirability of centralization and standardization, and GILS efficacy in supporting NIDR relative to other models.

## 7.0. PARTICIPANTS

Participants in the scripted online user assessments were selected from the University of North Texas and Syracuse University student bodies. Invitations and recruitment statements specified prerequisites of familiarity with the Web, Netscape, and fielded searching of online databases. The Texas group comprised four students; the Syracuse sessions involved six users. Table C5-1 summarizes the study group profile, characterized by means of the first eight questions on the script.

**Table C5-1**
**Scripted Online User Assessment—Participant Profile**

| **Background** |
|---|
| 1 "private citizen", 1 art undergraduate student, 1 political science undergraduate student, 1 history undergraduate student, and 6 library science graduate students |
| Average of more than 2 years' Internet usage |
| **Government Information Experience** |
| Print sources of government information, on average, searched monthly or less frequently |
| Frequency of searching online sources of government information varies from weekly to "as required by class" |
| Reports on government activity/public notices and legislation most frequently sought information |
| Only one participant had read, heard about, or used GILS (one encounter implied) |
| Most knew that Federal agencies have libraries |
| Half of group unaware of the function of purpose of many Federal agencies |
| Strong agreement that public electronic access to government information is important |
| **Searching Behavior** |
| Self-teaching through trial-and-error predominant method of acquiring/refining online searching skills |
| Browsing websites or bookshelves more common than use of online help, card catalogs, or application of professional training |
| One user claims 40% "success" in locating government information by starting with agency homepages |

## 8.0. SESSION DELIVERY: LOGISTICS AND EQUIPMENT

The two hosting universities employed different logistics for staging the script-execution and subsequent debriefing events. The University of North Texas online session was conducted in a graduate computer laboratory reserved for this special use; all users executed the script simultaneously; and all users met immediately upon completion of the online session for a 20-minute, relatively nondirective debriefing about GILS that also solicited additional feedback. At Syracuse, individual users executed the script serially over the course of 3 days at a dedicated workstation within an office and convened for a relatively traditional focus group of more than 1 hour duration early the following week. These variations in approach were not felt to degrade the quality of data but rather enhanced investigators' appreciation of the methods' flexibility and potential site-specific adaptation.

Baseline resources to conduct the sessions included workstations with:

- Pentium processor
- Netscape 2.0 or higher:
  - Options/Security Preferences set to disable security popups
  - Bookmarks set to script sites, or session homepage with links to script sites
- Well-behaved mouse
- Mousepad
- Clean screen
- Adequate seating
- Adequate lighting
- Adequate work area (recommended print orientation for script is landscape not portrait).

Other required materials included:

- (2) No. 2 sharp pencils with erasers
- Human subjects consent forms.

Prior to beginning the online session, users were asked to complete the demographic portion of the script and listen to a brief overview of session activities and expectations, including admonishments against deviating from the printed instructions and reassurance that research assistants were available to deal with the inevitable vagaries of the Internet such as error messages, site inaccessibility, etc. and to clarify script instructions.

## 9.0. METHOD LIMITATIONS

The exploratory technique of real-time scripted GILS user assessments presents several challenges to investigators and participants. First and foremost is development of a script that sieves out extraneous variables such as user interface and search engine functionality. Second, the script must be understandable by users having a wide variety of learning styles and reading comprehension levels. Third, the instrument must facilitate predictable and common searching "paces" among users with varying types and speeds of searching reflex. Last, the script must contain a narrow margin for both user- and system-induced errors.

In addition, participants must:

- be familiar with the Web and selected browser
- have at least rudimentary online search skills
- be capable of following instructions independently
- be capable of recording thoughts-in-process
- be willing to contribute up 2 hours' participation (if debriefing/focus group is included).

The method also demands a relatively high resource commitment in terms of standardly equipped computing workstations and in terms of human skill to create scripts, to coordinate session delivery, to lead and record focus groups, to analyze largely qualitative data, and to integrate those data into the larger picture of research objectives.

For users, the requisite constraint against "surfing" around the information space of GILS—as both a dedicated (self-contained) system and as an intellectual and/or virtual link to other resources—is unnatural and frustrating, and may lead to a perception that the speed and accuracy of mouseclicking is a hidden agenda of the research. It is important to explain to users that the tight scripting is critical to internal validity in terms of "investigators must be certain that they are assessing *specifically* and *only* what they set out to assess."

## 10.0. LESSONS LEARNED AND RECOMMENDATIONS TO FUTURE RESEARCHERS

During the development, delivery, and assessment of the scripted user online-assessment method, investigators identified the following areas for improvement.

### Script Development

Investigators found that a deliberate review of GILS record content (such as that performed in this evaluation study) not only informs the specification of evaluative criteria and selection of supporting scripted encounters but helps focus the script creator on *service output* characteristics rather than system-dependent variables such as interface and search engine performance. In addition, pretesting of the script as close to actual delivery time as possible is essential to clarify wording, optimize organization of the material, time the session, and confirm hyperlinks.

### Facility Readiness

Every attempt should be made to ensure the availability and adequacy of required resources (specified above) in order to maximize online time. The "required resource" of network response time, however, proved difficult to manage in the current study. (University of North Texas delivered a 6:00 p.m. session that featured instances of poor response comparable to that at Syracuse sessions run at 11:30 a.m.). Control or isolation of this variable is desirable to increase the reliability or internal validity of the script (i.e., to ensure that the user is responding to targeted GILS features); however, a naturalistic study of GILS should account for the effects of response time on user satisfaction.

### Supporting/Validating Activities

The results of a session such as those described above are obviously limited in generalizability—a valid cross-sampling of potential GILS user populations is most likely impossible, the script as instrument will inevitably carry the stamp of its creator(s) perception of GILS, and time does not permit adequate comparison by users of the multiple GILS approaches and output. Given these inherent limitations to the method, investigators strongly recommend that online assessments be supplemented (or "triangulated") with other user-oriented research methods such as a focus group, a library "site visit" for unobstrusive observation, "talk-aloud protocol" free-form search sessions, online user-satisfaction surveys, panel studies, and personal interviews.

## 11.0. CONCLUSION

The use of a scripted online user assessment of GILS proved complementary to the study's other evaluation methods and provided invaluable data about users' first impressions of current agency implementations. Advantages of the technique include recency of data, its applicability to a potentially diverse base of GILS users—from librarians and other intermediaries, to targeted searchers (e.g., students completing assignments), browsers of information (e.g., subject-oriented novices), and public-access advocates concerned with accountability and scope of coverage.

In addition, user-assessment scripts may be tailored specifically to a number of GILS implementation characteristics—e.g., current and planned search features, record presentation specifications, and depth/breadth of resource coverage—during critical stages in service maturity.

Data generated from actual users, especially real-time data, are essential in creating user-responsive, realistic objectives for GILS and in evaluating the service's performance to those objectives. It is hoped by the investigators that development and deployment of user-satisfaction instruments, such as this scripted assessment, will be recognized by agencies as worthy of the considerable investment required, and that methodologies will be shared in the spirit of continuous improvement.

**Attachment C5-1**

## INTRODUCTION AND ORIENTATION TO ONLINE SCRIPTED USER ASSESSMENT SESSION

Introduce oneself and other research team member(s) if present.

Paraphrase the following information when addressing study participants:

> This session will last about 1 hour and 15 minutes and be divided into roughly three parts. First, we will ask you to fill out a brief questionnaire about your online searching background. Then, for most of the session, you will use a set of written instructions to navigate through a Federal government information locator service. We'll talk about that more in a moment. Last, we will conduct a 15-minute debriefing in the conference room across the hall. This will give us a chance to listen to your impressions about the search session and the locator service in a relaxed, informal setting. If you haven't yet signed the consent form, please do so now.
>
> The written instructions you will follow to assess the information system are designed to draw out your opinions, impressions, and reactions—***not to test your knowledge about the government or about the World Wide Web or about database design***. With this in mind, please perform each exercise at a determined and positive rate rather than leisurely rate that you might normally use when searching on the web. We realize that this guideline imposes some unnatural restrictions on your curiosity, but after you perform the written instructions you will have 5 or 10 minutes to explore the system in a free-form fashion. Please take each question at face value and record your reactions honestly. Don't concern yourself with complete sentences or grammar, and feel free to use common abbreviations. Don't censor yourself! Again, we need your initial responses from the point of view of a consumer of a service.
>
> We emphasize that it is essential that you follow the instructions exactly, clicking on icons and hypertext only when asked. This not only saves time but ensures that each of you is experiencing the same feedback from the system. Also, do not look ahead at questions or go back to change completed answers; take one at a time. However, if at any time you find the instructions unclear or incorrect, or if the information on your screen seems out of sync with the instructions, please raise your hand right away, and one of us will get things back on track. We will occasionally look over your shoulder to see where you are in the instructions and may ask that you try to answer the questions more quickly.
>
> At about the midpoint of the written instructions, you will find our request that you take a short break. Please use this time to clear your mind of the system for a few minutes by visiting the restroom or water fountain directly across the hall or standing up to stretch and refocus your eyes. We don't want anyone to experience information overload!
>
> When you have answered all the questions in your instruction booklet, please hand it to one of us and head toward the snacks waiting in the conference room around the right-hand corner from the restrooms.
>
> Relax and have fun with this—take this opportunity to tell system designers what you really think of a service! Your responses are completely confidential. We feel very fortunate to have willing and interested volunteers, and *thank you* for taking this time to share your opinions with us.

*Note to researcher: Script Step 7 calls for your assistance if user retrieves "Helpful Hints" as one of three hits. Direct him/her to re-do the search with the term (with quotes)* ***"native american" OR indian***.